DOCUMENT RESUME

ED 227 797 HE 016 051

AUTHOR Korb, Roslyn
TITLE Clusters of Colleges and Universities: An Empirically Determined System.
PUB DATE [82]
NOTE 19p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Classification; *Cluster Analysis; *Colleges; Comparative Analysis; Higher Education; *Institutional Characteristics; *Universities
IDENTIFIERS *Peer Institutions

ABSTRACT

A technique for classifying higher education institutions was developed in order to identify homogenous subsets of institutions and to compare an institution with its empirically determined peers. The majority of the data were obtained from a 4-year longitudinal file that merged the finance, faculty, enrollment, and institutional characteristics survey of the Higher Education General Information Survey (1975-1978). The methodology included factor analysis of 189 variables and cluster analysis of institutions using information on the number, type, and levels of degrees awarded. The major doctoral institutions/research clustered into three major groups and two singlets. The 82 major doctoral institutions/non-research clustered into seven groups, four of which were singlets; the 287 comprehensive universities clustered into 10 groups, with 5 being singlets. Finally, the 686 general baccalaureate institutions clustered into 12 groups, with 4 being singlets, and 2 groups having 3 or fewer institutions. Appended materials include descriptions of the following eight dimensions that were used to cluster institutions: tuition dependence, instructional emphasis, black student enrollment, facilities, research emphasis, growth, size, and endowment level. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED227797

Clusters of Colleges and Universities: An Empirically Determined System

by
Roslyn Korb, Technical Planning Officer
Division of Postsecondary & Vocational Education Statistics, National Center for Education Statistics

Introduction

Traditional classification systems of higher education institutions, such as the Carnegie classifications or the newer NCHEMS classifications, separate institutions into very broad, summary categories that in many aspects have as much diversity within classes as between. As a result, little is gained in making comparisons among such broadly defined sectors or in using these classifications in micro-level analyses of individuals. On the other hand, it is not feasible to compare institutions on a one-to-one basis, nor would such an analysis be particularly relevant to the higher education community. Additionally, in looking at individual data in relation to higher education experiences or outcomes, previous work in this area suggests that there is sufficient noise in singular institutional characteristics to obscure any significant findings.

The purpose of the study was to identify higher-order dimensions of institutional characteristics that would permit homogeneous subsets of institutions to be identified. These subsets could then be used to compare an institution with its empirically determined peers. It would also provide school determinants with sufficient reliability such that it might be possible to detect interactions between higher education institutional characteristics and student characteristics.

The Data

Data for the study came from a four-year longitudinal file that was developed for NCES. This file merged the finance, faculty, enrollment, and institutional characteristics survey of HEGIS for 1975 through 1978. Additionally, several data elements from the ACE Title III files and BEOGS program data files were included. A summary of the variables and their source is contained in Table 1. All finance variables were adjusted for size using full-time equivalent enrollment.

The Methodology

Since it was felt that many of the HEGIS variables would be highly correlated, 189 variables representing each of the domains listed in Table 1 were factor analyzed using an alpha factor analytic technique. The resulting independent dimensions were rotated to

HE016 051

simple structure using the varimax criteria, and each institution was scored on each of the resulting factors, for use in subsequent analyses.

In defining homogeneous subsets of institutions, the analysis was limited to four-year schools. In addition, because the number and types of degrees awarded, and the programmatic emphasis of the institution were institutional dimensions that had not been included in the factor analysis, the clustering of institutions was carried out within categories representing the number, type, and levels of degrees awarded.

In particular, separate cluster analyses were done for major doctoral/research institutions, major doctoral/non-research institutions, comprehensive institutions and general baccalaureate institutions. Clusters of institutions were developed using the Fastclus subroutine of the Statistical Analysis System. This subroutine forms a cluster by examining the linear distance of every element to be clustered (in this case institutions) from a pre-defined cluster center, and assigning an element to a particular cluster on the basis of that distance. Since the number of clusters is defined a priori, the computational algorithm functions within the pre-defined number of clusters, rather than selecting an optimal number of clusters. Obtaining the optimal number of clusters is an empirical process, although by looking at the maximum distance within a cluster relative to the distance between cluster centers, it is possible to achieve a good fit to the data. For each of the four categories of institutions, the same factors were used to cluster institutions.

As a measure of the goodness of fit of the cluster solution to the data, discriminant analyses were performed. Since one objective of discriminant analysis is to determine the probability of group membership, a posteriori, it was felt that if all schools within a given cluster had a high probability of being in that cluster and a low probability of being in all other clusters, then the number of clusters could be considered sufficient, although not necessarily optimal.

The Results

The factor analysis resulted in 27 factors with eigenvalues greater than one. These accounted for 90% of the variability in the correlation matrix. Since many of these factors either accounted for a very small proportion of variance or were not relevant to the 4-year institution being considered, eight factors, accounting for 42% of the variance, were used in the cluster analysis. These eight factors are described in Table 2.

In considering these factors, it is interesting to note that 2 factors - tuition dependence and endowment - represent sources of institutional support, 2 represent institutional mission -

instruction and research, and 2 represent different dimensions of size - size of physical plant and number of students.

The factors were used in separate cluster analyses for the 4 institutional types and the results of these analyses will be discussed separately.

Major Doctoral Institutions/Research

These institutions clustered into three major clusters and two singlets. Singlets generally represent either very distinct institutions or extreme anomalies in the data. Table 3 presents the mean scores of the five clusters on each of the eight dimensions. For schools of this type, tuition dependence, instructional emphasis, research emphasis, and endowment levels are important dimensions. Although clusters were defined on the basis of eight independent dimensions, it is possible to get a graphic perspective of the results of the cluster analysis by looking at plots in 2-dimensional space. Figure 1 illustrates the five clusters of major doctoral institutions that are research-oriented for the tuition dependence and instructional emphasis dimensions. As may be seen, even on just two dimensions there is excellent separation between clusters, and integrity withing clusters.

Table 4 presents the results of the discriminant analysis procedure - indicating that each of the major clusters were sufficient, whereas singlets were placed in the cluster that they were closest to.

Major Doctoral Institutions/Non-research

Clustering this set of 82 institutions resulted in seven clusters, 4 of which were singlets. Table 5 contains the mean score of each cluster on each of the eight dimensions. For the three major clusters, tuition dependence, instructional emphasis, and endowment level seem to be the critical dimensions of difference, although the facilities dimension is also important in separating clusters A and C. A two-dimensional pictorial representation of the separation of these clusters is presented in Figure 2 - a plot of the endowment level and tuition dependence dimensions. The results of the discriminant analysis for these seven clusters are available in Table 6.

Comprehensive Universities

The 287 institutions in this class clustered into 10 groups, with 5 being singlets. This means of the clusters on the eight dimensions, shown in Table 7, indicate that tuition dependence separated all five primary clusters, while instructional emphasis distinguished clusters C and I from H. black student enrollment

distinguished G from the other 4 major clusters, facilities separated cluster H from cluster C, G, and I, and endowment level separated J and H from clusters C, G and I. The two-dimensional plot of endowment level by instructional emphasis shows some ambiguity among the defined major clusters, and suggests the importance of multidimensional criteria for clustering institutions. This ambiguity in two dimensions notwithstanding, the results of the discriminant analyses shown in Table 8 indicate the sufficiency of 5 major clusters for fitting these institutions.

General Baccalaureate Institutions

The 686 institutions in this class clustered into 12 groups, with 4 being singlets, and 2 groups having 3 or fewer institutions. The means of the 12 clusters on the eight dimensions are shown in Table 9. Among the six major clusters, tuition dependence, black student enrollment, and endowment seem to be the critical dimensions in separating general baccalaureate institutions. Two 2-dimensional plots actually illustrate how the multidimensional criteria for clustering operates. Figure 4 plots black student enrollment and tuition dependence. It can be seen in this plot how black students enrollment differentiates between schools in cluster D and all other clusters, while tuition dependence separates cluster E, L, and H. These two dimensions do not function well in separating clusters H and K, however. Figure 5 plots endowment level by tuition dependence. Here it may be seen that endowment separates cluster H and K, but cluster D is indistinct from cluster E. The result of the discriminant analysis are presented in Table 10. These discriminant analysis results are not quite as confirming as those achieved with the other three classes of institutions, suggesting either that more clusters might be appropriate, or that the dimensions on which these clusters were based were not as relevant to this class of institutions as they were for the three other institutional types.

Summary

The results of this study clearly demonstrate the viability of developing homogeneous subsets of institutions that can be used for comparing an institution with its empirically determined peers. Yet, the results also have methodological implications that should be considered in making practical use of this technique.

The results for comprehensive and general baccalaureate institutions indicated a need for using multiple dimensions that are also relevant to particular types of institutions if homogeneous clusters of institutions are to be developed and if definitive separation among clusters is to be achieved.

The results also suggest that the composition of a particular cluster is somewhat dependent on the dimensions used for clustering - that is, if different dimensions were used, institutions would

cluster in slightly different ways. Thus, in using this technique, consideration should be given to the comparisons to be made among peer institutions so that the dimensions used for clustering are consonant with the desired comparisons.

Table 1. Domains of variables used in the analysis and source.

Variable Domain	Source
Current funds revenues by source	HEGIS Finance Survey - Part A, 1978
Current Fund Expenditures and Mandatory Transfers	HEGIS Finance Survey - Part B, 1978
Physical Plant Assets	HEGIS Finance Survey - Part C, 1978
Indebtedness on Physical Plant	HEGIS Finance Survey - Part D, 1978
Endowment Assets	HEGIS Finance Survey - Part E, 1978
Changes in Fund Balances - Net Increases/Decreases	HEGIS Finance Survey - Part F, 1978
Number of Faculty	HEGIS Employee Survey, 1978
Ethnicity of Students	HEGIS Enrollment Survey, 1978
Number of Students	HEGIS Enrollment Survey, 1978
Tuition	HEGIS Institutional Characteristics Survey, 1978
Admission Requirements	HEGIS Institutional Characteristics Survey, 1978
Number of BEOG's Awards, Amount of Awards, and Title III Institution	ACE Title III File

Table 2. Description of Eight Dimensions Used to Cluster Institutions

Factor	Name	Description
I	Tuition Dependence	Schools having a high score on this factor both spend and earn money on auxiliary enterprises such as bookstores, athletics, dormitories, and food services. A high proportion of the revenue of these schools comes from student tuition and tuition charges for both undergraduate and graduate students are high. Additionally, schools with a high score on this factor tend to provide scholarships or grants for students, have a small number of part-time students, and a high number of faculty for the number of students in the school.
II	Instructional Emphasis	Schools with a high score on this factor spend a disproportionate amount on instruction and services that support instruction - such as audio/visual services, computing support, and course and curriculum development, as well as libraries, general administrative services, and research. These schools tend to receive revenues from State, Federal and private sources which are earmarked either for specific research projects of instructional or public service programs.
III	Black Student Enrollment	Schools with a high score on this factor tend to have a large number of students who receive substantial Basic Educational Opportunity Grant assistance. In addition, they have a high proportion of black undergraduate and graduate students and tend to award institutional grants to students.
IV	Facilities	Schools with a high score on this factor either have a very large campus or exist in an area where building and land costs are high. In order to maintain these facilities, shcools with a high score on this factor spend a disproportionate amount directly on the maintenance and operation of their buildings and land, and in ancillary services such as general administration and community relations.
V	Research Emphasis	Schools with a high score on this factor spend and earn revenues from major Federally funded research and development centers that are independent of the institution's primary mission. They also tend to emphasize research within the institution and probably add equipment in support of this research function. In addition to Federal sources of revenues, these schools receive money for specific projects through private gifts and endowment income.

Table 2. Continued

Factor	Name	Description
VI	Growth	Schools with a high score on this factor have had a large increase in funds available for the renewal and replacement of their physical plant and have apparently used these funds to add buildings to their physical plant. These schools have also acquired additional equipment, probably in support of instruction and public service functions.
VII	Size	Schools with a high score on this factor are large and have an even larger proportion of graduate students. They tend to receive money from State appropriations and pay higher than average faculty salaries.
VIII	Endowment Level	Schools having a high score on this factor would tend to have very large endowments with attendant high earnings, and realized income. Additionally, they would be building their endowments at a fairly high rate, and maybe channeling endowment income into the support of libraries, scholarships and services that support the institution's primary missions of instruction, research, and/or public services.

Table 3. Results of the cluster analysis for major Doctoral Institutions/research.

CLUSTER SUMMARY

CLUSTER	MEMBERS	MAX DISTANCE FROM CENTER
A	49	24.18821
B	10	12.93672
C	9	18.53726
D	1	0
E	1	0

CLUSTER CENTERS

CLUSTER	Tuition Depend	Instruc Emph.	Blk Stud Enr	Facilities	Research Emph	Growth	Size	Endowmen
A	8.73646816	3.78139686	-2.12572811	8.97079165	3.12519849	3.92793523	8.19423814	1.851965
B	16.39653165	14.31533839	-3.96978662	7.59874799	12.39996668	6.41383620	6.84016283	16.335841
C	28.92550327	28.49288845	-6.24404630	15.32766908	29.57382547	11.44743333	8.90352231	48.494002
D	21.78967071	37.23314575	-4.66534627	17.66724973	28.92962721	28.49345608	7.60082215	20.150907
E	39.94132995	54.02669265	-10.11168826	36.59717237	161.96147306	28.79641989	10.38120897	86.859128

Table 4. Number and Percent of Major Doctoral/Research Institutions Classified into Five Clusters by Discriminant Analysis.

Classified from Cluster	Classified into Cluster A		B		C		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
A	48	98.96	1	2.04	0	0.00	49	100.00
B	0	0.00	10	100.00	0	0.00	10	100.00
C	0	0.00	0	0.00	9	100.00	9	100.00
D	0	0.00	1	100.00	0	0.00	1	100.00
E	0	0.00	0	0.00	1	100.00	1	100.00

Figure 1. Plot of instructional emphasis by tuition dependence for Major Doctoral Institutions/Research

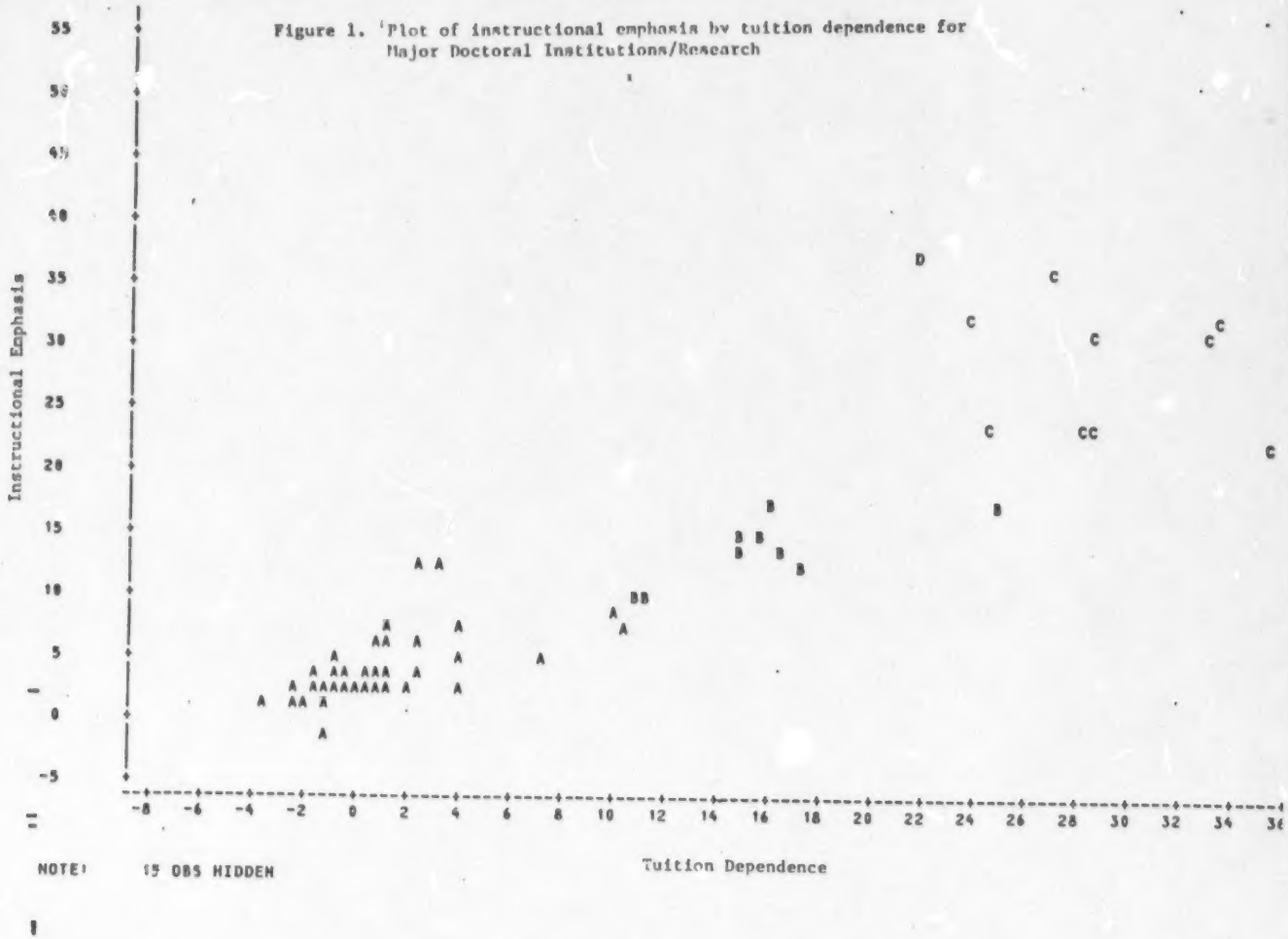

Table 5. Results of the cluster analysis for Major Doctoral Institutions/Non-research.

CLUSTER SUMMARY

CLUSTER	MEMBERS	MAX DISTANCE FROM CENTER
A	5	7.276152
B	1	0
C	58	13.67907
D	1	0
E	15	7.030889
F	1	0
G	1	0

CLUSTER CENTERS

CLUSTER	Tuition Depend	Instruc Emph	Blk Stud Enr	Facilities	Research Emph	Growth	Size	Endowment
A	16.047780352	18.707351708	-3.773265507	6.531761299	7.821139942	3.035495294	2.572774700	16.67959988
B	22.842809882	11.687342413	-6.927870033	9.410060408	14.170694317	5.022860448	2.646167272	47.28364720
C	-0.396556476	1.062604834	-1.427575221	-0.305585233	0.666429247	1.331317866	3.988229198	-0.61938375
D	-3.626591403	1.908547975	-2.616775952	-0.632911768	0.477478419	2.803246681	30.711117680	-1.93072920
E	0.017500907	4.910537191	-3.462787064	2.532431859	3.478251949	1.527802269	2.471916686	7.03960108
F	27.449929048	17.756859872	-5.505009107	10.757516596	13.632979635	5.044076083	3.870321385	31.38182997
G	5.931223478	11.789278180	18.169866994	5.929405616	6.231070616	10.328669010	5.844316365	3.56138141

Table 6. Number and Percent of Major Doctoral/Non-research Institutions Classified into Seven Clusters by Discriminant Analysis.

Classified from Cluster.	Classified into Cluster A		C		E		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
A	5	100.00	0	0.00	0	0.00	5	100.00
B	1	100.00	0	0.00	0	0.00	1	100.00
C	0	0.00	58	100.00	0	0.00	58	100.00
D	0	0.00	1	100.00	0	0.00	1	100.00
E	0	0.00	0	0.00	15	100.00	15	100.00
F	1	100.00	0	0.00	0	0.00	1	100.00
G	0	0.00	1	100.00	0	0.00	1	100.00

Figure 2. Plot of Endowment Level by Tuition Dependence for Major Doctoral Institutions/Non-research.

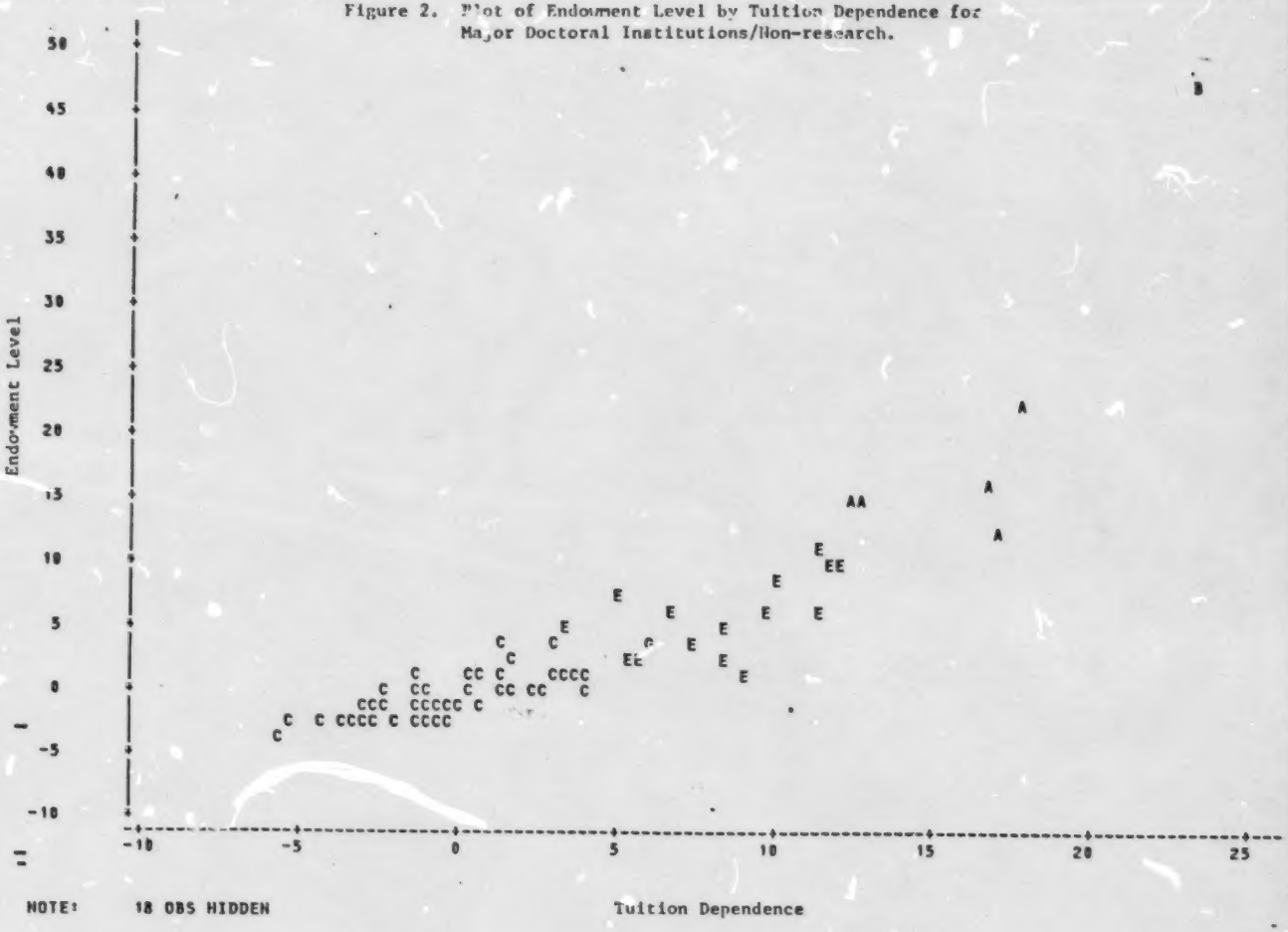

Table 7. Results of the cluster analysis for Comprehensive Universities.

CLUSTER SUMMARY

CLUSTER	MEMBERS	MAX DISTANCE FROM CENTER
A	1	0
B	1	0
C	180	11.87323
D	1	0
E	1	0
F	1	0
G	17	10.93935
H	3	5.419793
I	70	10.46208
J	12	9.859171

CLUSTER CENTERS

CLUSTER	Tuition Depend	Instruc Emph	Blk Stud Enr	Facilities	Research Emph	Growth	Size	Endowment
A	2.34888642	6.99529262	-1.25577998	3.81638214	9.17566497	9.97348195	2.96377661	6.11878085
B	12.37511174	15.74908589	-3.05519575	6.27131899	14.89156798	15.42581886	5.30540888	19.20812108
C	-1.99675036	-0.94184523	-0.37807591	-1.24617936	-0.68217636	-0.18257377	1.06406169	-1.85761998
D	-5.71662756	1.86661289	-0.37988487	-2.73468525	-0.39360954	3.05131471	18.28590230	-2.56497450
E	75.23143693	70.38992599	-10.24160486	59.43395237	49.32434914	33.88487848	9.53635357	57.82797927
F	-5.00078145	0.84722890	-2.77851158	-2.58785814	0.10208828	1.60518623	36.59562736	-2.68349647
G	-0.81317244	2.40897813	13.16039186	-0.20523238	0.42325898	0.97586821	1.58634748	-0.42855307
H	19.88548258	8.60379818	-4.96568223	8.02133758	6.[illegible]5816[illegible]	2.68343272	[illegible]23236[illegible]	27.76216533
I	4.94897267	1.34953916	-1.56791806	0.57213991	0.57[illegible]1153	-0.16122603	0.[illegible]77510	1.73748765
J	10.29517629	4.61955854	-3.18462214	3.48878574	3.52484588	1.86995752	0.[illegible]91746	12.54186138

Table 8. Number and Percent of Comprehensive Institutions Classified into Ten Clusters by Discriminant Analysis.

Classified from Cluster	Classified into Cluster C		G		H		I		J		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
A	0	0.00	0	0.00	0	0.00	0	0.00	1	100.00	1	100.00
B	0	0.00	0	0.00	0	0.00	0	0.00	1	100.00	1	100.00
C	179	99.44	1	0.56	0	0.00	0	0.00	0	0.00	180	100.00
D	1	100.00	0	0.00	0	0.00	0	0.00	0	0.00	1	100.00
E	0	0.00	0	0.00	1	100.00	0	0.00	0	0.00	1	100.00
F	1	100.00	0	0.00	0	0.00	0	0.00	0	0.00	1	100.00
G	0	0.00	17	100.00	0	0.00	0	0.00	0	0.00	17	100.00
H	0	0.00	0	0.00	3	100.00	0	0.00	0	0.00	3	100.00
I	0	0.00	0	0.00	0	0.00	70	100.00	0	0.00	70	100.00
J	0	0.00	0	0.00	0	0.00	2	16.67	10	83.33	12	100.00

Figure 3. Plot of Endowment Level by Instructional Emphasis for Comprehensive Institutions.

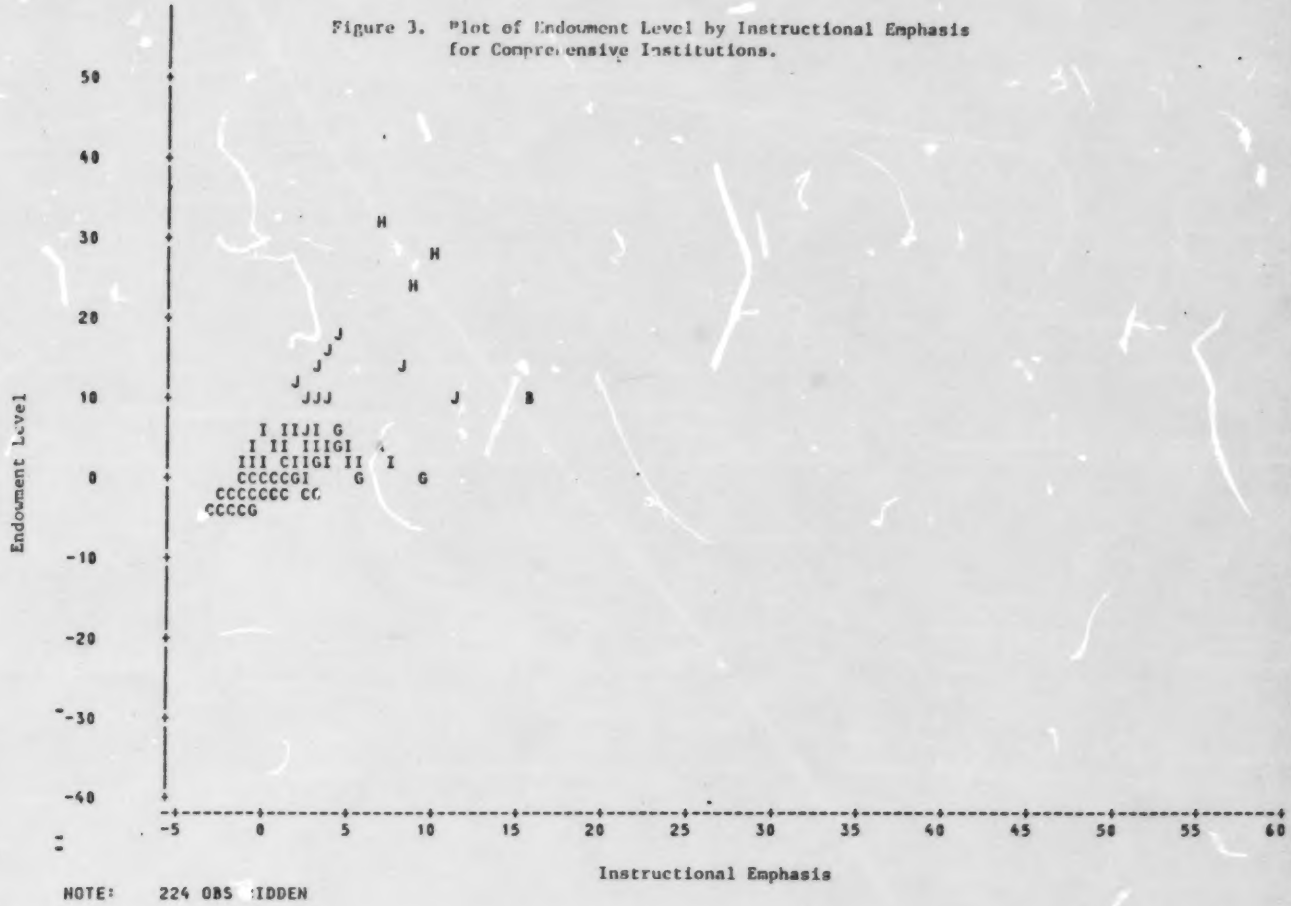

Table 9. Results of the cluster analysis for General Baccalaureate Institutions.

CLUSTER SUMMARY

CLUSTER	MEMBERS	MAX DISTANCE FROM CENTER
A	6	12.78529
B	1	0
C	1	0
D	57	13.84402
E	317	12.68437
F	1	0
G	1	0
H	10	12.91003
I	2	3.239072
J	3	7.711985
K	26	15.77647
L	261	10.54792

CLUSTER CENTERS

CLUSTER	Tuition Depend	Instruc Emph.	Blk Stud Enr	Facilities	Research Emph.	Growth	Size	Endowment
A	21.869573487	8.330150367	-5.029389511	9.102736989	8.766168398	3.207625188	7.436397277	39.59727703
B	7.907345581	7.550824932	-7.210257953	90.58733[illegible]883	4.078953290	4.712077246	-2.078881631	7.93710269
C	6.071611842	4.576508591	-3.558086931	6.168[illegible]936[illegible]1	5.149480975	1.746827156	-0.394[illegible]4704	28.3293[illegible]245
D	1.6[illegible]3792927	0.408858568	11.770901874	-0.525[illegible]208[illegible]8	-0.278001452	-0.405405784	-1.185187485	-0.746322456
E	-0.[illegible]29445914	-1.199637368	0.286417017	-0.729979[illegible]09	-0.850751109	-0.[illegible]25092587	-0.759234565	-1.328[illegible]4980
F	46.520235058	90.618580556	-2.647497849	18.614898[illegible]59	19.887375261	24.018792241	3.091298716	15.44777[illegible]498
G	11.938663742	4.259111305	-6.047208365	8.582366[illegible]67	4.101945695	-1.907947099	-0.845519024	56.63211[illegible]220
H	16.126844362	8.823662067	-0.113822162	9.857940[illegible]65	2.177570593	7.008104407	-0.616106052	5.458238085
I	11.683195281	7.230224430	-5.063110453	23.18627524[illegible]1	5.315326113	2.510685281	0.045700175	22.352844355
J	5.221428566	4.875636666	-2.612587518	21.1746664[illegible]5	1.281054238	1.833622710	-1.090733606	3.956264105
K	15.644415938	5.389378150	-2.985706781	5.888668505	4.538349174	1.629041970	-0.282613508	19.353552611
L	6.167598121	0.876139026	-0.471980747	1.534897014	0.296630183	0.074569873	-0.7[illegible]45864	2.582395515

Table 10. Number and Percent of General Baccalaureate Institutions Classified into Twelve Clusters by Discriminant Analysis.

Classified from Cluster	Classified into Cluster A		D		E		H		I		J		K		L		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
A	5	83.33	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	16.67	0	0.00	6	100.00
B	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	100.00	0	0.00	0	0.00	1	100.00
C	1	100.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	100.00
D	0	0.00	55	96.49	2	3.51	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	57	100.00
E	0	0.00	0	0.00	311	98.11	0	0.00	0	0.00	0	0.00	0	0.00	6	1.89	317	100.00
F	0	0.00	0	0.00	0	0.00	1	100.00	0	0.00	0	0.00	0	0.00	0	0.00	1	100.00
G	1	100.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	100.00
H	0	0.00	0	0.00	0	0.00	9	90.00	0	0.00	0	0.00	0	0.00	1	10.00	10	100.00
I	0	0.00	0	0.00	0	0.00	0	0.00	2	100.00	0	0.00	0	0.00	0	0.00	2	100.00
J	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	3	100.00	0	0.00	0	0.00	3	100.00
K	2	7.69	2	7.85	0	0.00	0	0.00	0	0.00	0	0.00	22	84.62	1	3.85	26	100.00
L	0	0.00	0	0.00	12	4.60	2	0.77	0	0.00	0	0.00	3	1.15	244	93.49	261	100.00

Figure 4. Plot of Black Student Enrollment by Tuition Dependence for General Baccalaureate Institutions.

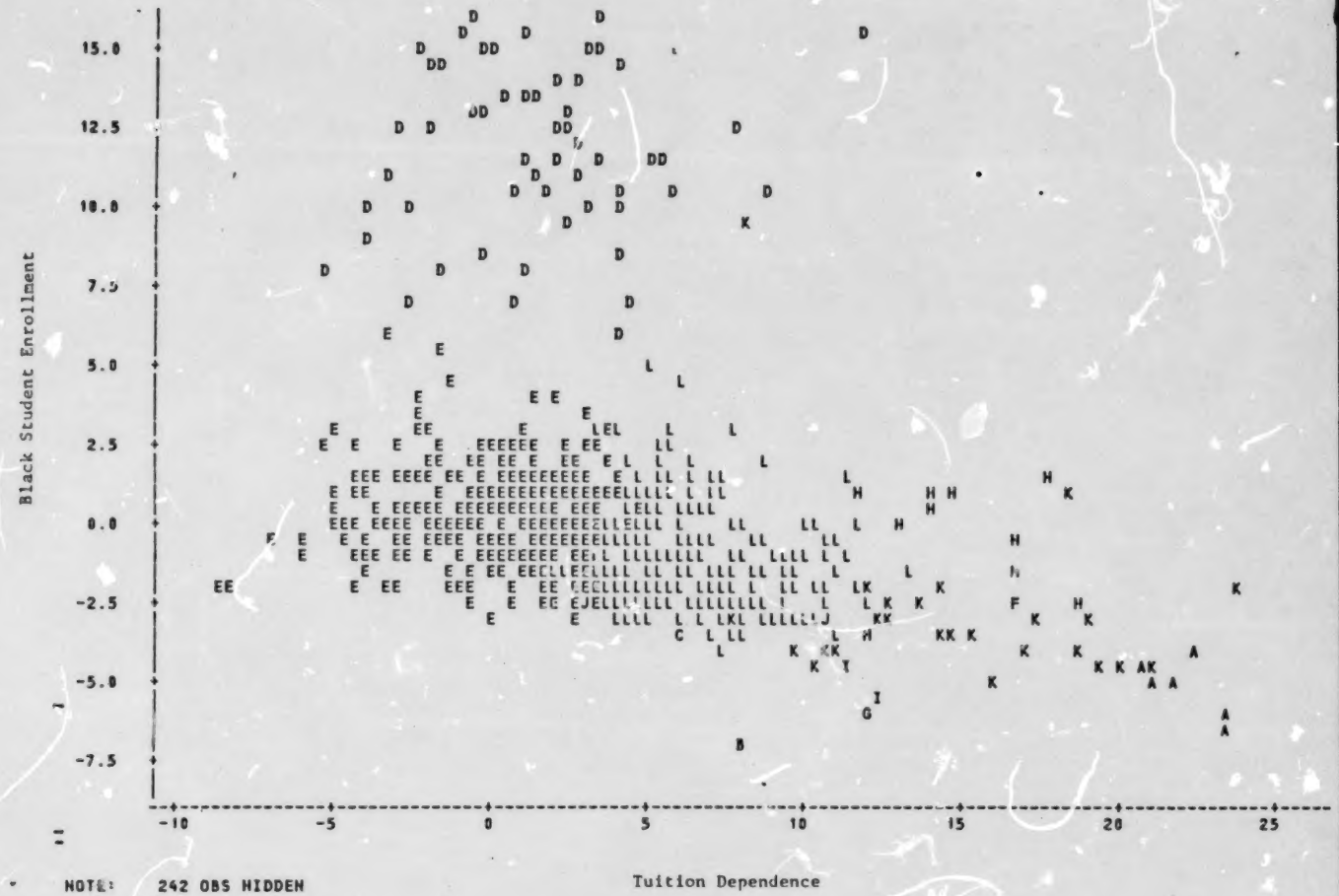

Figure 5. Plot of Endowment Level by Tuition Dependence for General Baccalaureate Institutions.